

README for the Reproducibility Package of the analytical output of the manuscript “Improving the Quality of Survey Estimates from Longitudinal Studies: An Application to the LSMS Panel Surveys”

Contents

1. Overview
2. Data Availability
3. Instructions for Replicators
4. List of Exhibits
5. Software Requirements

Overview

This reproducibility package (RP) reproduces the analytical output (estimators, tables, and graphs) of the manuscript “Improving the Quality of Survey Estimates from Longitudinal Studies: An Application to the LSMS Panel Surveys.” The RP directory at the OneDrive link [Reproducibility_package](#) comprises the data used as inputs in the analysis, a set of programs in Stata and R, a series of temporary datasets, and the final outputs in the form of tables and graphs, which are part of the final manuscript. The latter, to be submitted to the World Bank Policy Research Working Papers Series, is an integral part of the package as well, being not published yet. All data used in this work are publicly available. A master Stata dofile “MASTER.do” will run the entire set of programs, comprising the R routines thanks to a function which recalls the R program in Stata. **The estimated time for running the master file is 2 minutes.**

Data Availability

- All data are publicly available, although the versions used in this work may not necessarily correspond to those online. Details on the data sources are found in the paragraph below.

Data Sources

- **Uganda National Panel Survey 2005-2009** – datasets used in this work are the following:

Year	Datasets	Source	Link to the data	Version	Download date	Note
2005	GSEC1, GSEC2, GSEC18	World Bank Microdata library	https://microdata.worldbank.org/index.php/catalog/1001	UGA_2005_UNPS_v01_M_STATA.zip	March 2018	Data currently available on the microdata library are V2

2009	GSEC1, GSEC2 GSEC3	World Bank Microdata library	https://microdata.worldbank.org/index.php/catalog/1001	UGA_2009_UNPS_v01_M_STATA.zip	March 2018	Data currently available on the microdata library are V2
2009	pov2009_10	World Bank Microdata library	https://microdata.worldbank.org/index.php/catalog/1001	UGA_2005_2009_UNPS_v02_M_STATA8.zip	October, 2024	

- **National Panel Survey 2010-2011** – Datasets GSEC2, available on the World Bank Microdata library at <https://microdata.worldbank.org/index.php/catalog/2166>, downloaded on October, 2024.
- **National Panel Survey 2011-2012, Wave III** – Datasets GSEC2, available on the World Bank Microdata library at <https://microdata.worldbank.org/index.php/catalog/6743>, downloaded on October, 2024.
- **National Panel Survey 2013-2014** – Datasets GSEC1, GSEC2, GSEC3, GSEC19, UNPS 2013-14 Consumption Aggregate, available on the World Bank Microdata library at <https://microdata.worldbank.org/index.php/catalog/2663> , downloaded on October, 2024.
- **National Panel Survey 2015-2016** – Datasets gsec1, gsec2, gsec3, gsec19, pov2015_16 Consumption Aggregate, available on the World Bank Microdata library at <https://microdata.worldbank.org/index.php/catalog/3460> , downloaded on October, 2024.
- **Uganda_2010** – Datasets hh_final.csv, Individual_final.csv – available on the Food and Agriculture Organization of the United Nations, RuLIS - Rural Livelihoods Information System - Household-based surveys page at <https://www.fao.org/in-action/rural-livelihoods-dataset-rulis/data-application/bulk-download/derived-micro-variables/en> , downloaded on December 2020.
- **Uganda_2014** – Datasets hh_final.csv, Individual_final.csv – available on the Food and Agriculture Organization of the United Nations, RuLIS - Rural Livelihoods Information System - Household-based surveys page at <https://www.fao.org/in-action/rural-livelihoods-dataset-rulis/data-application/bulk-download/derived-micro-variables/en> , downloaded on December 2020.
- **Uganda_2016** – Datasets hh_final.csv, Individual_final.csv – available on the Food and Agriculture Organization of the United Nations, RuLIS - Rural Livelihoods Information System - Household-based surveys page at <https://www.fao.org/in-action/rural-livelihoods-dataset-rulis/data-application/bulk-download/derived-micro-variables/en> , downloaded on December 2020.

- **04_2019National_Population_Projections_by_5_year_age_groups_and_sex.xlsx**, 2015 Male and Female, Column A,B,C, available on Uganda Bureau of Statistics, Population and Censuses web page, at https://www.ubos.org/wp-content/uploads/publications/04_2019National_Population_Projections_by_5_year_age_groups_and_sex.xlsx, downloaded on October 2024. Required data were saved in a txt format to be used in R.

Instructions for Replicators

New users should follow these steps to run the package successfully:

1. The first step in this reproducibility package is to download the needed sources, please follow the links in the data source section above. The downloaded datasets will need to replace what you now find in the 1_Input_data_folder. Please, note that the version of the **Uganda National Panel Survey 2005-2009** used in this work is not the one currently online, and therefore this is included in the package for consistency.
2. You must update the directory paths in the MASTER.do dofile and the R path in the main R script. The Master.do will run all the STATA and R routines used in this analysis.
3. Make sure all required software and dependencies are installed as listed in the [Requirements](#) section. These are already included in the ado folder and renv.lock
4. Run the MASTER.do dofile.

List of Exhibits

While one tables is replicated from and external source, all the other tables can be reproduced running the codes in this RP. Final graph and tables are generated running the **D_Sankey_diagram_panel.do** and **C2_tables.do**. For both these two files to run, however, you must generate temporary datasets running the codes which:

1. prepare the data (**A1_Paneldataset.do, A2_ind_hh_characteristics.do, A3_including_Rulis_variables.do**), and
2. generate the estimators (**B1_Load_Uganda_data.R, B2_Computation Base_weight_GWSM.R, B3 Household transformation.R, B4_Computation GWS 2013.R, B4_Computation GWS 2015.R, B5 Creation auxiliary variables.R, B6 Creation target variables.R, B7 Calibration 2015.R, B8 Calibration 2015 with hhs.R**).

The table below maps all tables and figure in the paper to its generative script:

Exhibit name	Output filename	Script	Note
Figure 6.1	sankey.gph sankey.jpg	0_Sankey_diagram_panel.do – line 92	Found in 4_Output

Exhibit name	Output filename	Script	Note
Table 6.2	04_2019National_Population_Projections_by_5_year_age_groups_and_sex.xlsx, available at https://www.ubos.org/wp-content/uploads/publications/04_2019National_Population_Projections_by_5_year_age_groups_and_sex.xlsx , downloaded on October 2024	NA	Found in 1_Input_data
Table 6.3	Statistics.xlsx, Sheet: S1	C2_tables.do lines 25-155	Found in 4_Output
Table 6.4	Statistics.xlsx, Sheet: Employed	C2_tables.do lines 391-441	Found in 4_Output
Table 6.5	Statistics.xlsx, Sheet: Pov	C2_tables.do lines 391-444	Found in 4_Output
Table 6.6	Statistics.xlsx, Sheet: S2 – Line 1	C2_tables.do lines 173 - 326	Found in 4_Output
Table 6.7	Statistics.xlsx, Sheet: rur_urb – Line 1	C2_tables.do lines 326 - 378	Found in 4_Output
Table 6.8	Statistics.xlsx, Sheet: S2 – Line 10	C2_tables.do	Found in 4_Output
Table 6.9	Statistics.xlsx, Sheet: rur_urb – Line 15	C2_tables.do lines 326 - 378	Found in 4_Output

Software Requirements

To run this RP the following software and dependencies are required:

- **Stata - version 18**
 - sankey v18
 - palettes
 - rsource

- **R - version 4.4.0**
 - ReGenesees 2.4
 - varhandle
 - dplyr
 - devtools
-